**MSplusR: An R Package for Visualization**

**of MS/MS Spectral Data Using UMAP**

**Hiroyuki Yamamoto**

[h.yama2396@gmail.com](mailto:h.yama2396@gmail.com)

Japan Computational Mass Spectrometry (JCompMS) group

**Abstract**

Mass spectrometry (MS) data analysis encompasses complex processes including preprocessing, data filtering, similarity matrix computation, dimensionality reduction, and visualization. In this study, we introduce **MSplusR**, an R package designed to streamline these workflows and enhance reproducibility in metabolomics, proteomics, and lipidomics research. The package provides functionalities for efficient spectrum data processing, computation of similarity matrices based on shared spectral peaks, and advanced dimensionality reduction methods such as UMAP. Additionally, **MSplusR** includes an interactive visualization tool leveraging R Shiny and Plotly, which allows researchers to explore UMAP projections alongside corresponding MS/MS spectra. This paper outlines the implementation details, use cases, and applications of **MSplusR** in high-throughput MS data analysis, providing a robust and user-friendly framework for researchers.

**Keywords:** mass spectrometry, metabolomics, R package, UMAP, spectrum analysis, reproducible research.

**MSplusR: UMAPを用いたMS/MSスペクトルデータ可視化のためのRパッケージ**

質量分析インフォマティクス研究会　山本 博之　h.yama2396@gmail.com

**Abstract**

質量分析データ解析は、前処理、データフィルタリング、類似度行列の計算、次元削減、可視化を含む複雑なプロセスを必要とします。本研究では、これらの作業を効率化し、再現性を向上させるRパッケージ「MSplusR」を紹介します。このパッケージは、スペクトルデータの効率的な処理、共有スペクトルピークに基づく類似度行列の計算、UMAPなどの高度な次元削減手法を提供します。さらに、R ShinyとPlotlyを活用したインタラクティブな可視化ツールを含み、研究者がUMAPの結果を対応するMS/MSスペクトルと共に探索できる環境を提供します。本稿では、MSplusRの実装詳細、使用例、質量分析データ解析への応用について説明し、研究者にとって信頼性が高く使いやすいフレームワークを提供します。

**Introduction**

質量分析（Mass Spectrometry, MS）は、代謝物、ペプチド、タンパク質などの分子を特定し定量するための分析技術として広く利用されています。メタボロミクスやプロテオミクスの急速な発展により、データセットの複雑性と規模が増大しています。これに伴い、効率的なデータ処理ツールが求められていますが、柔軟性や再現性、統合性に課題が残っています。

これらの課題に対処するために、質量分析データ解析の包括的かつ再現性のあるパイプラインを提供するオープンソースのRパッケージ「MSplusR」を開発しました。このパッケージは、堅牢な前処理機能、統計解析ツール、インタラクティブな可視化機能を組み合わせています。Rの広範なエコシステムを活用することで、既存のバイオインフォマティクスワークフローと互換性がありながら、柔軟性を維持しています。

**Method**

MSplusRは、質量分析データ解析を効率化するためのさまざまな機能を提供します。まず、mzML形式のスペクトルデータを読み込み、ピーク検出、ビニング、正規化といった前処理を行います。これにより、データは統一された形式で整理され、後続の解析が可能になります。また、スパース行列表現を利用して、共有スペクトルピークに基づく類似度行列を効率的に計算します。この類似度行列は、クラスタリングやネットワーク解析に不可欠な要素となります。

次元削減手法としては、UMAPやPCAを統合し、高次元データを低次元空間に可視化可能な形式で変換します。これにより、データ内のパターンや関係性を視覚的に理解できます。さらに、MSplusRにはR ShinyとPlotlyを活用したインタラクティブな可視化ツールが含まれており、研究者がUMAP結果と対応するMS/MSスペクトルを探索できる環境を提供します。このツールには履歴管理機能もあり、選択されたスペクトルの履歴を追跡し比較することが可能です。

**Application**

MSplusRは、質量分析研究のさまざまな分野で応用可能です。メタボロミクスでは、mzMLファイルを処理し、類似度行列を計算することで、代謝パターンの特定や類似サンプルのクラスタリングを支援します。プロテオミクスやリピドミクスの分野では、複雑なスペクトルを解析し、バイオマーカーの特定やサンプル間の関係性の視覚化を効果的に行えます。

さらに、インタラクティブビューアは、データ品質の検証や特定のスペクトル特徴の調査に役立ちます。このように、MSplusRは多様な分析ニーズに応える柔軟性と機能を備えています。

**Conclusion**

MSplusRは、質量分析データの前処理、類似度計算、次元削減、可視化を統合的に提供する堅牢で使いやすいフレームワークです。高度な統計手法とインタラクティブな可視化機能を組み合わせることで、メタボロミクスやプロテオミクス研究における重要な課題に対処します。今後の開発では、対応する解析の範囲を拡大し、計算効率を向上させることを目指しています。これにより、MSplusRは大規模質量分析データ解析における不可欠なツールとなるでしょう。

**公開情報**

MSplusRは、GitHub上でオープンソースRパッケージとして公開されています：<https://github.com/hiroyukiyamamoto/MSplusR>。詳細なドキュメントとサンプルワークフローが提供されており、研究での採用と再現性の確保を促進します。

グラフ, 散布図

自動的に生成された説明

